

①9 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

①1 N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

2 807 188

②1 N° d'enregistrement national : 00 04064

⑤1 Int Cl⁷ : G 06 T 15/70, G 10 L 13/00

⑫ DEMANDE DE BREVET D'INVENTION

A1

②2 Date de dépôt : 30.03.00.

③0 Priorité :

⑦1 Demandeur(s) : VRTV STUDIOS Société à responsabilité limitée — FR.

④3 Date de mise à la disposition du public de la demande : 05.10.01 Bulletin 01/40.

⑤6 Liste des documents cités dans le rapport de recherche préliminaire : Se reporter à la fin du présent fascicule

⑦2 Inventeur(s) : BOUJU BASTIEN, ARCHAMBEAUD JEROME, PERCHE JACQUES, DE ZORZI FREDERIC et MACEIRAS RAMON.

⑥0 Références à d'autres documents nationaux apparentés :

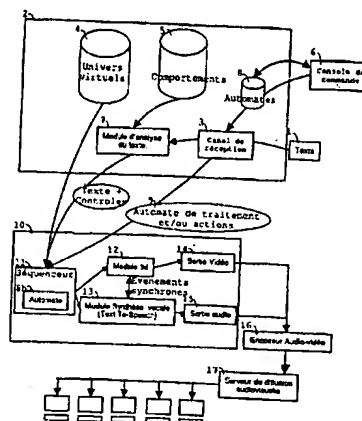
⑦3 Titulaire(s) :

⑦4 Mandataire(s) : BREESE MAJEROWICZ SIMONNOT.

⑤4 EQUIPEMENT POUR PRODUCTION AUTOMATIQUE EN TEMPS REEL DE SEQUENCES AUDIOVISUELLES VIRTUELLES A PARTIR D'UN MESSAGE TEXTUEL ET POUR LA DIFFUSION DE TELLES SEQUENCES.

⑤7 La présente invention se rapporte à un équipement pour la production automatique, en temps réel, de séquences audiovisuelles virtuelles à partir d'un message textuel exprimé dans un langage naturel et pour la diffusion de telles séquences, du type comportant un module de synthèse vocale (13), un module graphique (12) pour la génération d'images de synthèse et un processeur pour piloter les interfaces d'entrée et de sortie ainsi que lesdits modules, caractérisé en ce qu'il comporte une première mémoire dans laquelle est enregistrée une base de données de comportement (5), un module d'analyse de texte (7) recevant les messages provenant d'une interface (1) d'entrée et générant des marqueurs comportementaux issus de la base de données de comportements (5) en fonction des messages d'entrée, lesdits marqueurs pilotant le module graphique (12), l'équipement comportant en outre des moyens pour la synchronisation du module de synthèse vocale et du module graphique et des circuits audio et vidéo recevant les données provenant des modules de synthèse vocale et graphique pour délivrer un signal audiovisuel.

La présente invention se rapporte également à un procédé de production automatique, en temps réel, de séquences audiovisuelles virtuelles.



FR 2 807 188 - A1



**EQUIPEMENT POUR PRODUCTION AUTOMATIQUE EN TEMPS REEL DE
SEQUENCES AUDIOVISUELLES VIRTUELLES A PARTIR D'UN MESSAGE
TEXTUEL ET POUR LA DIFFUSION DE TELLES SEQUENCES.**

5 La présente invention concerne le domaine des équipements pour la production automatique de séquences audiovisuelles virtuelles à partir d'un message textuel et pour la diffusion de telles séquences.

10 On connaît dans l'état de la technique différentes solutions pour produire des séquences d'animation virtuelles associées à la synthèse vocale.

 Le brevet américain US 5 970 459 par exemple décrit un système pour la synchronisation entre une image animée et un convertisseur de synthèse vocale.

15 Le brevet US 5 630 017 décrit un équipement pour l'animation des mouvements de lèvres d'une image de synthèse en fonction de données textuelles.

20 Le but de l'invention est d'automatiser en temps réel la production d'images animées audiovisuelles dont le comportement est représentatif du contenu syntaxique et grammatical et du contexte des données textuelles provenant de sources extérieures et d'assurer leur diffusion en temps réel ou différé.

25 L'invention concerne selon son acception la plus générale un équipement pour la production automatique, en temps réel, de séquences audiovisuelles virtuelles à partir d'un message textuel exprimé dans un langage naturel et pour la diffusion de telles séquences, du type comportant un module de synthèse vocale, un module graphique pour la génération d'images de synthèse et un
30 processeur pour piloter les interfaces d'entrée et de sortie ainsi que lesdits modules, caractérisé en ce qu'il comporte une première mémoire dans laquelle est enregistrée une base de données de comportement, un module d'analyse de

texte recevant les messages provenant d'une interface d'entrée et générant des marqueurs comportementaux issus de la base de données de comportements en fonction des messages d'entrée, lesdits marqueurs pilotant le module graphique, l'équipement comportant en outre des moyens pour la synchronisation du module de synthèse vocale et du module graphique et des circuits audio et vidéo recevant les données provenant des modules de synthèse vocale et graphique pour délivrer un signal audiovisuel.

Les moyens et modules peuvent être indifféremment constitués de circuits électroniques ou de modules purement informatiques.

Le terme de signal audiovisuel se rapporte à tout type de signal permettant de réaliser au moins une image et du son, y compris, par exemple, à une image programmée.

Le langage naturel du message textuel s'entend de toute langue qui ne soit pas un langage informatique artificiel.

Avantageusement, l'équipement selon l'invention comporte en outre une base de données d'univers virtuels contenant des données numériques pour produire des décors et personnages animés, la base de données de comportements comportant des descripteurs pour le pilotage par le module d'analyse de texte et les animations comportementales de la base d'univers virtuels.

Avantageusement également, l'équipement selon l'invention permet de réaliser des animations en temps réel et de les diffuser en temps réel, voire en temps différé.

Avantageusement enfin, l'équipement selon l'invention permet d'éviter d'utiliser un langage informatique en entrée.

Selon une variante, il comporte une machine de rendu comportant un automate de traitement du texte dont les états contiennent les actions à réaliser sur les scènes provenant des bases de données.

5 Selon une autre variante avantageuse, il comporte un module de sortie pour numériser et compresser le signal audiovisuel, la sortie dudit module étant reliée à un serveur de diffusion sur un réseau de télécommunication, en temps réel ou en temps différé.

10 L'invention concerne également un procédé de production automatique, en temps réel, de séquences audiovisuelles virtuelles, caractérisé en ce que l'on procède à une analyse du texte provenant d'une interface d'entrée, pour produire des marqueurs comportementaux issus
15 d'une base de données de comportements en fonction du message d'entrée, lesdits marqueurs pilotant un module graphique synchronisé avec un module de synthèse vocale pour générer un signal audiovisuel.

20 L'invention sera mieux comprise à la lecture de la description qui suit, se référant à la figure 1 annexée représentant le schéma de principe d'un équipement selon un mode de réalisation non limitatif de l'invention.

La figure 1 représente l'architecture générale d'un équipement selon l'invention.

25 Une interface (1) reçoit un fichier texte, par exemple depuis un clavier de saisie, ou une base de données, ou encore par un réseau informatique, par exemple Internet, ou par un réseau téléphonique, par reconnaissance vocale ou analyse de la voix.

30 Un serveur (2) reçoit le fichier texte, par un canal de réception (3) piloté par une console de commande (6) assurant la définition du contexte de diffusion et le cas échéant la mise à jour en temps réel des données.

Le serveur comporte une base de données de comportement (5) contenant des marqueurs correspondant aux correspondances entre les unités lexicales et syntaxiques du texte et les animations comportementales de la base de données d'univers virtuels (4) contenant des scènes en trois dimensions, par exemple, au format vrml.

Un module d'analyse de texte (7) est chargé d'insérer dans le texte source des marqueurs comportementaux issus de la base de données (5).

Une machine de rendu (10) produit les scènes 3D animées et sonorisées, à partir d'un automate et du texte. La machine de rendu (10) comprend un automate de traitement (9). Les états de l'automate contiennent les actions sur les scènes (chargement, animation, ...) et des scripts. Les transitions entre les états se feront à partir des mots et marqueurs du texte issus du module d'analyse (7). La définition de l'univers virtuel à charger (décors, personnages) se trouve dans l'état initial de l'automate.

Le texte issu du module d'analyse (7) et contenant les mots et les marqueurs, est lu par le séquenceur (11) qui déclenche la prononciation du texte et les actions sur la scène 3D par l'intermédiaire de l'automate (8b). Le module graphique (12) en trois dimensions et le module de synthèse vocale (13) sont synchronisés en temps réel. Il est possible de déporter le module de synthèse vocale sur une autre machine.

Le module de synthèse vocale (13) génère le signal audio qui est traité par une interface de sortie (15). Le module graphique (12) génère le signal vidéo qui est traité par une interface de sortie (14). Un encodeur (16) numérise les données audio et vidéo et génère un signal numérique comprimé dirigé vers un serveur de diffusion (17).

L'ensemble des opérations, depuis l'entrée du
texte jusqu'à la génération des images et la synthèse du
son, est effectué en temps réel. La diffusion de masse peut
5 être effectuée en temps réel ou en temps différé, suivant
le procédé utilisé.

La machine de rendu (10) peut éventuellement
être déportée sur chaque terminal.

REVENDICATIONS

1. Equipement pour la production automatique,
en temps réel, de séquences audiovisuelles virtuelles à
5 partir d'un message textuel exprimé dans un langage naturel
et pour la diffusion de telles séquences, du type
comportant un module de synthèse vocale (13), un module
graphique (12) pour la génération d'images de synthèse et
un processeur pour piloter les interfaces d'entrée et de
10 sortie ainsi que lesdits modules, caractérisé en ce qu'il
comporte une première mémoire dans laquelle est enregistrée
une base de données de comportement (5), un module
d'analyse de texte (7) recevant les messages provenant
d'une interface (1) d'entrée et générant des marqueurs
15 comportementaux issus de la base de données de
comportements (5) en fonction des messages d'entrée,
lesdits marqueurs pilotant le module graphique (12),
l'équipement comportant en outre des moyens pour la
synchronisation du module de synthèse vocale et du module
20 graphique et des circuits audio et vidéo recevant les
données provenant des modules de synthèse vocale et
graphique pour délivrer un signal audiovisuel.

2. Equipement pour la production automatique,
en temps réel, de séquences audiovisuelles virtuelles à
25 partir d'un message textuel, selon la revendication 1
caractérisé en ce qu'il comporte en outre une base de
données d'univers virtuels (4) contenant des données
numériques pour produire des décors et personnages animés,
la base de données de comportements (5) comportant des
30 descripteurs pour le pilotage par le module d'analyse de
texte et les animations comportementales de la base
d'univers virtuels (4).

3. Equipement pour la production automatique,
en temps réel, de séquences audiovisuelles virtuelles à

partir d'un message textuel, selon la revendication 1 ou 2
caractérisé en ce qu'il comporte une machine de rendu (10)
comportant un automate de traitement (9) dont les états
contiennent les actions à réaliser sur les scènes provenant
des bases de données (4 et 5).

4. Equipement pour la production automatique,
en temps réel, de séquences audiovisuelles virtuelles à
partir d'un message textuel, selon l'une au moins des
revendications précédentes caractérisé en ce qu'il comporte
un module de sortie pour numériser et compresser le signal
audiovisuel, la sortie dudit module étant reliée à un
serveur de diffusion (17) sur un réseau de
télécommunication.

5. Procédé de production automatique, en temps
réel, de séquences audiovisuelles virtuelles, caractérisé
en ce que l'on procède à une analyse du texte provenant
d'une interface d'entrée, pour produire des marqueurs
comportementaux issus d'une base de données de
comportements (5) en fonction du message d'entrée, lesdits
marqueurs pilotant un module graphique synchronisé avec un
module de synthèse vocale pour générer un signal
audiovisuel.

Figure unique

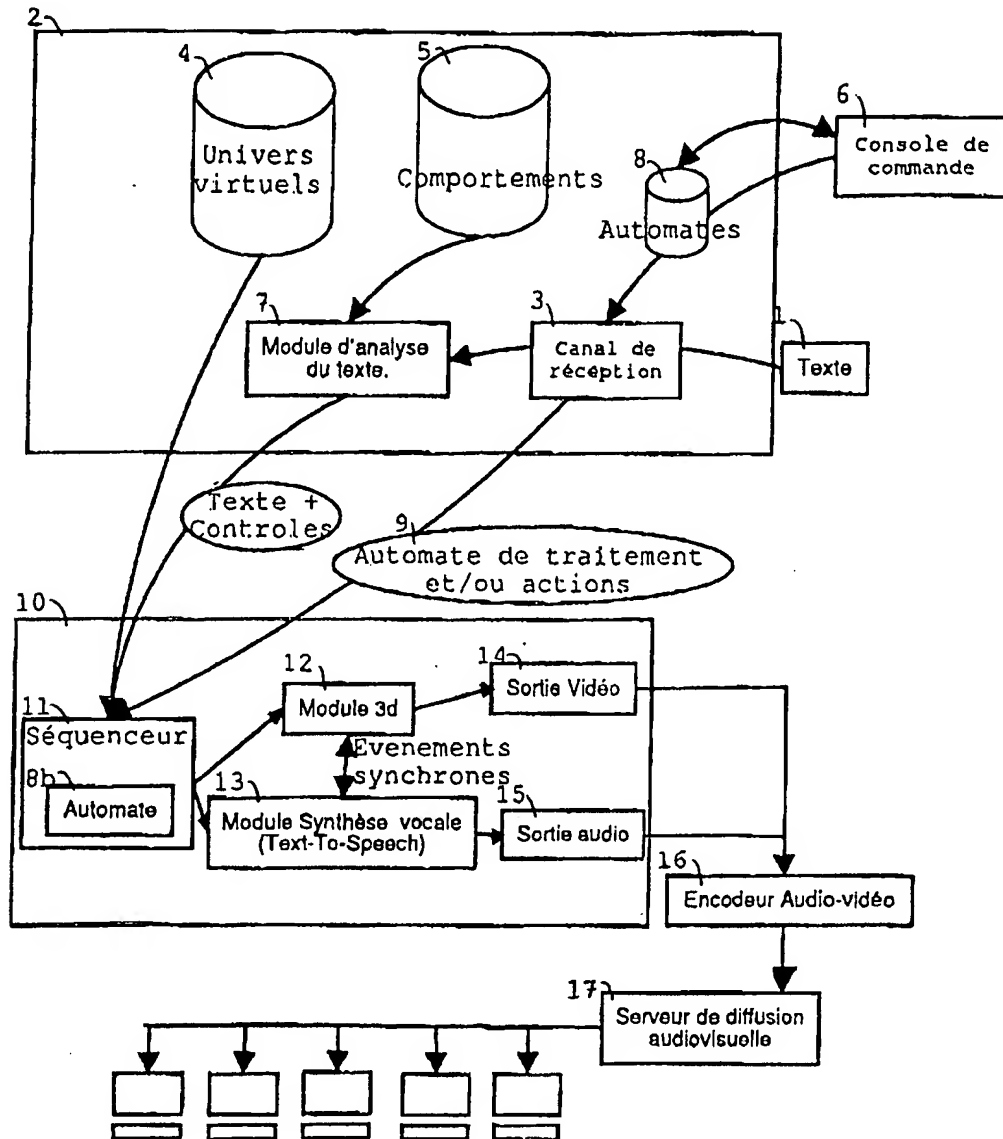


Figure 1



RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

2807188
N° d'enregistrement
national

FA 585053
FR 0004064

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	YAMADA A ET AL: "Visual text reader for virtual image communication on networks" 1997 IEEE FIRST WORKSHOP ON MULTIMEDIA SIGNAL PROCESSING (CAT. NO.97TH8256), PROCEEDINGS OF FIRST SIGNAL PROCESSING SOCIETY WORKSHOP ON MULTIMEDIA SIGNAL PROCESSING, PRINCETON, NJ, USA, 23-25 JUNE 1997, pages 495-500, XP002147180 1997, New York, NY, USA, IEEE, USA ISBN: 0-7803-3780-8 * le document en entier *	1-5	G06T15/70 G10L13/00
A	EP 0 598 597 A (CANON INFORMATION SYST INC) 25 mai 1994 (1994-05-25) * page 3, ligne 33 - ligne 49; figure 2 *	1-5	
A	PATENT ABSTRACTS OF JAPAN vol. 1995, no. 10, 30 novembre 1995 (1995-11-30) & JP 07 175909 A (CANON INC), 14 juillet 1995 (1995-07-14) * abrégé *	1,5	
A	EP 0 860 811 A (DIGITAL EQUIPMENT CORP) 26 août 1998 (1998-08-26) * abrégé *	1,5	
A	EP 0 896 322 A (AT & T CORP) 10 février 1999 (1999-02-10) * abrégé *	1,5	
			DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7)
			G10L
Date d'achèvement de la recherche		Examineur	
12 septembre 2000		Wanzeele, R	
CATÉGORIE DES DOCUMENTS CITÉS			
<p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p>			